

REMARKS

Claims 1-23 are all the claims presently pending in the application. Claims 1, 4-7, 12-17 and 20-23 have been amended to more clearly define the invention. Claims 1, 7, 17 and 23 are independent.

These amendments are made only to more particularly point out the invention for the Examiner and not for narrowing the scope of the claims or for any reason related to a statutory requirement for patentability.

Applicant also notes that, notwithstanding any claim amendments herein or later during prosecution, Applicant's intent is to encompass equivalents of all claim elements.

Claims 1-23 stand rejected under 35 U.S.C. § 102(e) as being anticipated by Marc Alexander Najork (U.S. Patent No. 6,321,265).

This rejection is respectfully traversed in the following discussion.

I. THE CLAIMED INVENTION

The claimed invention is directed to a method for searching files stored on a network.

The method includes accessing a first file on the network, downloading data from that first file and setting an access time to access a second file based on the downloaded data from the first file. The downloaded data provides an indication of when the second file is scheduled to be updated.

Conventional network file search engines conduct searches for updated files on networks periodically, such as at regular intervals. One problem with these conventional systems is that these systems do not have any method for determining when a website might be scheduled to be updated. Depending on how often a website is updated, the web crawler's archive data could be very outdated. On the other hand, frequent web crawler visits to websites which are not frequently updated consumes valuable computer resources.

The present invention provides a method for determining when and how often a web crawler should return to a web site. The present invention provides this advantage because the method accesses a first file on a network, downloads data from the first file and sets an access time to access a second file based upon the data from the first file, where that downloaded data indicates when the second file is scheduled to be updated.

In an exemplary embodiment of the present invention, the method accesses a channel definition file (CDF) which provides an indication of when a particular channel (and/or subchannel) is scheduled to be updated (see page 4, line 15 - page 5, line 2). Therefore, in this exemplary embodiment the first file is the CDF and the second file is the channel.

In this manner, the present invention provides for more efficient web crawling of a web site by crawling the site when and where it is likely the information contained therein is updated (page 6, lines 7-15).

II. THE PRIOR ART REJECTION

The Examiner alleges that the Najork et al. reference teaches the claimed invention. Applicant submits, however, that there are elements of the claimed invention which are neither taught nor suggested by this reference.

The Najork et al. reference discloses a web crawler which downloads data sets from a plurality of host computers. The Najork et al. reference is concerned with avoiding overloading a host computer with multiple parallel requests to download data from the same host computer. Overloading a host computer with multiple parallel requests can diminish that host computer's responsiveness to page requests or may even cause the host to crash (col. 1, line 60 - col. 2, line 2).

The Najork et al. reference does not teach or suggest setting an access time for a second file based on data from a first file. Rather, the Najork et al. reference discloses a system which sets an access time based upon the download time of a previous document from the same web server (col. 2, lines 43-45). In other words, the Najork et al. reference attempts to avoid multiple parallel requests to the same host computer by estimating how long a file being currently downloaded (second file) will take to download based upon the time that a previous document (first file) took to download and then setting an access time for a subsequent download (third

file) based upon that amount of time. In that manner, the system disclosed by the Najork et al. reference sets an access time for accessing the subsequent download (third file) based upon the download time of a document (first file) which is previous to a document (second file) currently being downloaded. Therefore, the Najork et al. reference does not teach or suggest setting an access time based upon data downloaded FROM a first file. Rather, the Najork et al. reference discloses setting an access time based upon a download time of a first file.

Indeed, the Najork et al. reference does not even address the problem solved by the present invention. The Najork et al. reference is directed to a current visit to a web site by a web crawler and downloading all of the data from the host computer of that web site. The Najork et al. reference addresses issues regarding avoiding overloading the host computer with multiple parallel requests during that same visit.

In stark contrast, the present invention is directed to determining when a web crawler should conduct a return visit to a host computer. The present invention is concerned with visiting a web site more often than necessary. As explained above, conventional web crawlers (including the web crawler disclosed by the Najork et al. reference) visit web sites periodically. The problem is that the data on each web site may not have been updated since the last visit. Therefore, these conventional web crawlers revisit these web sites too often.

The present invention is directed to determining when to conduct a return visit to a web site based upon data from that web site which may indicate when a file is scheduled to be updated. As explained above, in an exemplary embodiment of the present invention, the first file corresponds to a channel definition file which includes data about when a second file which

corresponds to a channel is scheduled to be updated. The present invention takes advantage of that data to determine when to conduct a return visit to download the claimed second file with the most update information.

By contrast, the Najork et al. reference is directed to a current visit to a web site where a first web page is downloaded and analyzed to retrieve addresses for additional web pages on the same host computer. Therefore, the Najork et al. reference is not at all concerned with when the first web page may be updated. Rather, the Najork et al. reference is concerned with when the host computer may safely download additional web pages.

Therefore, contrary to the allegations of the Examiner the Najork et al. reference does not teach or suggest each and every element of the claimed invention. Therefore, the Examiner is respectfully requested to withdraw this rejection of claims 1-23.

III. FORMAL MATTERS AND CONCLUSION

In view of the foregoing amendments and remarks, Applicant respectfully submits that claims 1-23, all the claims presently pending in the Application, are patentably distinct over the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Should the Examiner find the Application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary in a telephonic or personal interview.

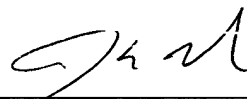
Serial No. 09/672,304
Docket No. AM9-99-0146

13

The Commissioner is hereby authorized to charge any deficiency in fees or to credit any overpayment in fees to Attorney's Deposit Account No. 50-0481.

Respectfully Submitted,

Date: 3/12/03



James E. Howard
Registration No. 39,715

McGinn & Gibb, PLLC
8321 Old Courthouse Rd., Suite 200
Vienna, Virginia 22182
(703) 761-4100
Customer No. 21254

Attachment:
Excess Claim Fee Payment Letter